DOCUMENT RESUME

ED 184 076 CS 005 321

AUTHOR Lucas, Peter A.
TITLE Visual Attention During Sentence Verification.
PUB DATE Apr 79
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 8-12, 1979). For related documents see CS 005 071, CS 005 321-323 and ED 178 861.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Students; *Eye Fixations; *Eye Movements; Grammar; Higher Education; *Reading Processes; *Reading Research; Semantics; *Sentences; *Syntax

ABSTRACT

Eye movement data were collected for 28 college students reading 32 sentences with sentence verification questions. The factors observed were target sentence voice (active/passive), probe voice, and correct response (true/false). Pairs of subjects received the same set of stimuli, but with agents and objects in the sentences reversed. As expected, sentence voice and probe voice interacted strongly, such that reaction times were faster when these voices matched. The majority of additional response time produced by voice mismatch was attributed to decision processes occurring after the initial reading of the sentence. The relatively long times typically spent on the second noun phrase of the probe sentences seemed to reflect post-perceptual cognitive manipulations of the semantic information from the two sentences. This result supported a weak version of the "immediate linguistic awareness" hypothesis, in which the sentence is read and encoded into some unspecified internal representation prior to the comparison process. In a two-stage model of this particular reading task, an active primary linguistic encoding stage transformed the visual stimulus into some internal representation that was placed in a buffer store until an appropriate time for further semantic processing. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED184076

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Visual Attention During Sentence Verification

Peter A. Lucas

University of Wisconsin--Madison

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Peter A. Lucas

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1979.

CS005324

The psychological study of reading and of language comprehension in general must necessarily address itself to two conceptually distinct issues: A theory of language use must ultimately specify both the processes and the products of comprehension. But while conceptually distinct, the two issues are in practice highly interdependent. Much recent work in the psychological study of comprehension has been flawed by a failure to adequately consider this interdependence when proposing models of linguistic processes. Specifically, the willingness on the part of researchers to make ad hoc assumptions about the cognitive products of comprehension and to ignore the process by which such products are produced has led to the development of process models which, although successful in accounting for the data which motivated them, are of questionable value when applied to the study of ongoing comprehension of communicative language in normal contexts.

In the 1960's when psycholinguistics was in large part oriented toward assessing the psychological usefulness of "standard" Chomskian transformational theory, the issues of process and product were most manifest in what came to be known as the "derivational theory of complexity" and the "coding hypothesis" respectively. The former referred to the proposition that sentence perception somehow involved the reverse application of the transformations of the standard transformational theory in order to derive

underlying "base forms" of sentences given their surface forms as input. Similarly, the coding hypothesis asserted that the nature of these base forms was isomorphic to the "deep structures" of the then standard linguistic theory. Both of these hypotheses are now widely regarded as having been disproved (Fodor, Bever, and Garrett, 1974; but c.f., Valian, 1979). Their passing can be viewed as marking the end of the first major era of psycholinguistic research, during which linguistic theory held hegemony over psycholingusitic inquiry.

Subsequent work in the field has tended toward an increased consideration of semantic and pragmatic aspects of language use, and less emphasis on purely syntactic factors. Another major influence during this period has come from the areas of artificial intelligence and computer simulation. In particular, psycholinguists have been attracted to propositional network notations, in which information is represented as a collection of nodes representing semantically primitive "concepts" connected together with links corresponding to basic semantic relations. From these and other factors have emerged the paradigm of sentence verification models which have become a strong candidate for the new "standard theory" of language comprehension.

Pioneered by Clark (1969; Clark and Chase, 1972) and by Trabasso (e.g., 1972), this paradigm relies heavily on data derived from chronometric studies involving the verification of the truth value of sentences (or answering questions)

with respect to some state of affairs (e.g., a picture or another sentence). Making the assumption of additive, serial processing in the tradition of Sternberg (1966), the aim has been to develop explicit models of the verification process expressed in terms of the matching and manipulation of abstract propositional representations of the linguistic information.

It is instructive to examine this class of theory in terms of the process/product distinction with which we began. It is clear that, for most theorists in this area, abstract semantic structures--represented either as nested strings of atomic propositions or as the kinds of semantic networks described above--have taken the place of the linguistic deep structures of the coding hypothesis as the formalism used to characterize the internal representations which are the products of the comprehension process. The appropriateness of such representations have been treated as almost axiomatic by many theorists.

With respect to the processes of comprehension, matters are somewhat less clear. In one sense, this aspect of the problem is not addressed at all. The flowcharts traditionally used to describe the verification models typically begin with a box labelled "encode sentence," and few theorists in this tradition have ventured much speculation as to the inner workings of this box. In other words, the process of parsing sentences into propositions is seen as a unitary, automatic first step in the verification

process which is regarded as outside the domain of inquiry. In another sense, however, these are very much process models since they are intended to characterize the explicit cognitive processes involved in verification tasks. There is some controversy as to whether such performance tasks ought to be considered part of "comprehension" or rather as post-comprehension processes, but this distinction need not detain us here.

It should be pointed out that common to both the early transformational theorists and the current propositional approach is an implicit assumption which Levelt (1977) has called the "immediate linguistic awareness" (ILA) hypothesis. According to this hypothesis, the initial stage of all tasks involving sentence comprehension includes a common set of operations which result in the derivation of some underlying internal representation of the semantic content of the sentence.

The theoretical advantages of such an assumption are obvious. Among other things, it permits a convienent modular approach to theory building. The complexities of language-specific syntactic operations can be assigned to the parser which--until a model of these processes is forthcoming--the less linguistically inclined psychologist can treat as a black box out of which emerges propositions to serve as input for information processing models of subsequent "semantic" processes. Additionally, the ILA hypothesis is consistent with the introspective evidence

that sentences seem to be "understood" at some level automatically and without effort whether or not any subsequent use is made of the information acquired. For these and other reasons, the ILA hypothesis has been assumed, explicitly or implicitly, by a majority of workers in the field. A major issue of the research to be reported here is the extent to which this assumption, and the models which rely on it, are appropriate to the characterization of comprehension as it occurs both in sentence verification tasks and in normal reading.

Difficulties With the Sentence Verification Paradigm

Despite its popularity and success in accounting for experimental data, there are a number of criticisms that can be leveled at the class of models which has emerged form this research tradition. Tanenhaus, Carroll and Bever (1976) have criticized the ad hoc nature of the proposed specific forms of the propositional representations. Such representations are often chosen arbitrarily so as to simplify the construction of a model to fit the data of a given task domain, and without external motivation. A good fit of the data to the model is then offered as evidence for the psychological reality of both the process model and the representations upon which it operates. As Anderson (1976) and others have argued, there is an indeterminacy introduced by the fact that neither the assumed internal structures nor the operations hypothesized to operate on them are directly

observable. That is, given an arbitrary change in the assumed representation, a corresponding change to the process model can often be found which would yield predictions identical to the original (Anderson has made the claim that this indeterminacy is fundimental and essentially inescapable given the state of the art in cognitive psychology, but his argument is based upon a rather narrow set of assumptions which are subject of debate; c.f., Hayes-Roth, 1979; Pylyshyn, 1979). Consequently, in the absence of external motivation for a particular representational format, or at least convergence across studies, it seems unwise to put much faith in either the proposed models or in the assumed representations on which they depend.

Another matter of concern with the paragigm is the extent to which it in fact addresses issues of the comprehension process as it occurs in normal language use. We have already mentioned the problem of the task-specificity apparent in most of the models. A more fundimental issue is the appropriateness of assuming the ILA hypothesis in normal language usage. As we have seen, the bulk of the variance in verification time data produced by manipulation of syntactic forms is attributed in the models to post ILA symbol matching and recoding processes. The case can be made that this is less a well-motivated theoretical claim than an artifact of the preponderance of experimental procedures employing successive rather than

simultaneous measures--notably verification times--in developing and testing the models. Such techniques are incapable of differentiating initial comprehension effects from subsequent processing. This, combined with the aforementioned practical advantages of assuming ILA create a strong bias in favor of adopting it.

On the other extreme, if one were to consider comprehension as a process of acquiring information from the text or utterance rather that one of building some semantic internal representation of the text, one might be tempted to drop the assumption of an obligatory, purely syntactic initial parser entirely. From this view, it might seem more plausible to assume that syntactic, semantic, conceptual, and pragmatic variables all have their primary effect on the initial perception of the sentence. The patterns of response latency found in verification tasks may be less a reflection of post-perceptual "proposition matching" operations than of the differential difficulty of perceiving different kinds of sentences in different contexts. Such a position is compatible with arguments concerning the importance of semantic factors in syntactic processes starting at the earliest stages of the parsing process (e.g., McCawley, 1968; Olson, 1970). Furthermore, it is a characteristic of many of the more successful artificial intelligence efforts in natural language understanding that the programs include a simultaneous application of syntactic and semantic constraints (e.g., Schank, 1975; Winograd,

1972).

Other shortcomings of the sentence processing paradigm become apparent when one asks what the models have to say about language use in normal communicative contexts, or more specifically, what they might contribute to a theory of reading. When considered for this purpose, it is apparent that the near universal assumption of a manditory initial propositional encoding step (which is never analyzed as an explicit part of the model) severly restricts their ability to make interesting statements about these more complex processes. Since the aspects of the verification process represented in the models concern task-specific operations performed after an initial encoding, it is not obvious how they might contribute to a theory whose primary goal presumably concerns the construction of such representations. It one were to assume the naive view that reading consisted simply of constructing propositional representations of each input sentence, an extrapolation from the models would produce a flowchart consisting merely of a string of "encode sentence" boxes, and little more. Viewed this way, the models--even if correct--seem largely irrelevent to reading.

A somewhat more realistic view of discourse comprehension, however, might characterize the process not as one of simply encoding sentences into discrete propositions, but rather of forming an integrated semantic representation of the passage as a unit. Such a view has

gained wide acceptance in large part as a result of the work of Bransford and his colleagues (e.g., Bransford, Barclay and Franks, 1972; Bransford and Franks, 1971). Within such a framework, a major burden of a theory of comprehension is an explication of the processes by which incoming information is coordinated with and appended to the existing partial representation at any given point. It is possible that proposition-matching operations similar to the kind postulated in the verification models may be invoked in this kind of processing. In this way, the sentence-processing work might contribute, at least by analogy, to an understanding of the comprehension of extended text.

Clark and Haviland (1977) have taken a step in this direction in their analysis of patterns of given/new information in conversational text. Starting with Grice's (1967) notion of the Cooperative Principle and his conversational maxims, they present an insightful analysis of various devices by which speakers cooperate with listeners in their effort to map new linguistic information onto existing memory structures. The discussion is presented in terms of the same general kind of propositional representations that Clark has employed elsewhere. It does not appear, however, that the analysis is in any way dependent on this particular formalism--the process of given/new coordination seems equally necessary given any kind of internal representation that assumes integration at acquisition time. Thus, while the analysis is persuasive,

it seems to provide no strong motivation for any particular representation. However, a next logical step for a researcher in this tradition might be to propose that the specific processes involved in matching new propositions to old structures is identical to that postulated in the sentence verification models. Such a position would make quite specific predictions about the instantaneous processing load while reading sentences of various linguistic constructions embedded in texts with different patterns of given/new information. If such a position were to prove viable, the existing sentence processing literature might be applicable in a quite straightforward manner to discourse comprehension.

To summarize the discussion thus far of our original question of how models of sentence processing might contribute to the study of prose comprehension, we have examined two possibilities: (1) that the phenomena studied in these experiments are "post-comprehension" in nature and too task-specific to be of relevance to a theory of reading; and (2) that models resembling those proposed for sentence processing will prove to usefully describe the process of constructing integrated internal representations out of the fragmented propositional content of the input sentences. It should be noted that either of these possibilities can be accepted without violating the assumptions of the paradigm, notably a proposition-like internal representation of the linguistic information and the ILA hypothesis. If one is

willing to forgo these assumptions--particularly the latter--then a third possibility becomes apparent. It is possible that the models themselves are misleading or irrelevent to an understanding of prose processes but that the empirical phenomena studied may nonetheless prove important in some other theoretical context.

It is not my intention to discuss such an alternative theoretical context here. However, in very general terms, it might be supposed that the previously acquired semantic information (i.e., the linguistic context) affects the difficulty of the initial perceptual encoding of a sentence, rather than merely post-perceptual semantic operations. Such a view is compatible with certain weaker forms of the hypothesis testing model of comprehension and with the notion that words serve an informational function (in Shannon's sense) by specifying meaning relative to a set of linguistically or extra-linguistically defined alternatives (Olson, 1970; 1972). Thus, the increase in response time observed when verifying sentences of different syntactic forms may reflect not a comparison of mismatched propositions, but rather, a difficulty in processing a syntactic form inappropriate to the particular set of alternatives being considered by the reader/listener, thus disrupting the extraction of information from the text.

The important point is that in this view, it is the words themselves (i.e., the surface form of the sentence) that are mapped onto the existing internal representation,

and not some "deep" propositional representation derived on a sentence-by-sentence or clause-by-clause basis.

It is noteworthy that one of the early studies using a sentence-picture verification task was motivated by just such a view of the comprehension process. Olson and Filby (1972) introduced their report of a set of studies on verifying active and passive sentences with the following statement:

> The general hypothesis underlying this series of experiments was that when sentences are contextually appropriate, that is, when they match the structure of alternatives being considered by the reader...they are comprehended in terms of their surface grammatical structure.

Later in the paper, however, they presented a verification model which, while not actually based on a propositional representation, still included the ubiquitous "code sentence" stage, thus seemingly endorsing the ILA hypothesis in some form. Apparently, the authors had in mind the initial construction of some representation which preserved the surface structure of the sentence. This is perhaps the weakest form of the ILA hypothesis, since in the extreme case the internal representation may be considered as an unanalyzed sensory trace, which, presumably, must itself serve as input to the linguistic processor.

In practice, the ILA hypothesis cannot be empirically distinguished from the direct comprehension position using

the traditional successive measurement techniques (e.g., reaction time or error rate) employed in verification tasks. Several experimenters (e.g., Glucksberg, Trabasso, & Wald, 1973) have attempted to separate "comprehension time" from "verification time," typically by having subjects push a button when they understood the sentence and before the onset of the verification task. This technique can only be applied, however, to the encoding of the first stimulus in a verification task and thus does not really address the effects of context on comprehension. More relevant are studies by Dooling (1972) who measured subjects' reported comprehension time for sentences presented in varying degrees of context. He found that increased context produced faster judgements of meaningfulness of the sentences. However, when the task required judging whether the sentence fit into the provided context, "comprehension" times were slower. These results suggest that Dooling's subjects applied a more stringent criterion for "comprehension" in the high context conditions. This effect would seem to place severe restrictions on the subject self-report technique as a measure of comprehension time.

Eye Movements as a Measure of Comprehension Time

It seems clear that if the ILA hypothesis is to be experimentally evaluated under conditions of normal language processing, then some form of simultaneous measurement of processing load is required. In the present research, the monitoring of eye movements during reading was employed for

this purpose. Although this seems a rather natural choice of measures, there have been surprisingly few attempts at using eye movement techniques to address psycholinguistic issues. In fact, in his otherwise remarkably thorough review of the sentence perception literature, Levelt (1977) did not even include eye movements in his list of simultaneous measurement techniques. The technical difficulty of collecting eye movement data has no doubt contributed to the lack of enthusiasm for the technique. More fundamental, however, is the fact that, despite a long history of often heroic efforts at the application of eye movement techniques to reading research, the net yield of this effort in terms of theoretical or practical knowledge is widely regarded as being small (e.g., Tinker, 1958). The vast majority of this research, however, has been directed toward the study of global properties of eye movements per se, often seeking, for example, to distinguish good and poor readers on the basis of their characteristic eye movements.

There has recently been something of a renaissance of interest in eye movement research, partly because new technologies have broadened the range of available techniques. Most recent work, however, has focused on rather low-level perceptual factors such as the control processes involved in eye movements (e.g., Haber, 1976; Rayner and McConkie, 1976) or the size and nature of the perceptual span (McConkie and Rayner, 1975; 1976). The few studies which have addressed issues of language processing

have typically sought to demonstrate that eye movements reflect surface syntactic properties of sentences (Mehler, Bever, and Carey, 1967; O'Regan, 1975; Rayner, 1977; Wanat, 1971). The only researchers working within the sentence verification paradigm who have employed eye movement techniques (as far as I am aware) have been Just and Carpenter. Much of their reported work has been limited to studying eye movements during the verification of single sentences or sentence fragments (Just and Carpenter, 1976). More relevant, however, are a series of recent studies (Carpenter and Just, 1977) which used eye movements to investigate integrative processes in reading connected text. They were able to demonstrate longer total fixation time when the old/new structure of sentences didn't match that of their linguistic context. However, the subjects in these studies were required to search the text for inconsistencies while reading, a task which may have contributed to the effects. Furthermore, the old/new manipulations were effected by means of the cleft and pseudocleft constructions, producing rather anomolous-sounding texts in the mismatched conditions.

In the present experiment, eye movements were monitored during a straightforward sentence/sentence verification task. In this task, readers were presented with a simple sentence in either the active or the passive voice (e.g., "The bus driver greeted the passengers."/"The passengers were greeted by the bus driver.") followed immediately by a

yes/no question which either matched or mismatched the sentence with respect to voice ("Did the bus driver greet the passengers?"/"Were the passengers greeted by the bus driver?"). The subjects were required to answer the question by pressing one of two buttons. Previous experiments of this type (e.g., Anderson, 1974; Garrod and Trabasso, 1973; Olson and Filby, 1972) have consistently found a strong interaction between sentence voice and question voice such that immediate verification latencies were far shorter when the voices matched. Within the tradition of the proposition-matching models discussed above, this effect has almost always been assumed to reflect post-ILA symbol matching operations and the manipulation of a "truth index" whose value is changed each time a constituent mismatches. Eye movement data were collected to make it possible to identify the locus of the extra response time and thus to distinguish this model from one which assigns the increased response time to a slower initial comprehension of the question. Specifically, the latter hypothesis would predict longer and/or more forward fixations during the initial reading of the probe sentence under the voice mismatch condition. The ILA hypothesis would predict one of two effects: either (1) the bulk of the extra time occurs at or after the final fixation on the probe and before the response is made; or (2) The extra time is consumed during regressive movements after the initial reading of the probe. The latter effect would be predicted

if it is assumed that the eye tends to spontaneously fixate on the current symbol being processed (data of Just & Carpenter, 1976, in a somewhat similar task seem to reflect such an effect). In either case, there is no reason to predict longer initial reading times for mismatched sentences given the assumption of ILA.

Method

Subjects. Twenty-eight undergraduate volunteers served as subjects. Each had responded to advertising during summer session at Cornell University and was paid for participating. All had normal, uncorrected vision. Each subject had previously participated in at least one warmup session and so was familiar with the eye monitoring apparatus. None, however, had previously participated in a sentence verification experiment. Three subjects' data were lost because of experimenter error.

Materials and design. The stimuli consisted of 32 single sentences of the form

Subject Noun Phrase + Verb Phrase + Object Noun Pharse

In all cases, the length of both noun phrases and the active form of the verb phrase was exactly 10 characters (including spaces). The passive form of each verb phrase was 17 characters long. Each sentence was reversible in the sense that both noun phrases were potent and could sensibly serve as agent of the sentence. Another list of 32 sentences was prepared for use as practice items. This list was similiar

to the experimental list but was not constrained in constituent length. On a given trial, each target sentence was followed with an associated probe sentence which was identical to it except that on half the trials, it differed in voice. Additionally, on half the trials the agent and object of the probe sentence was reversed producing "false" trials.

The design of the study included three within-subject factors: target sentence voice (active/passive); probe voice (active/passive); and correct response (true/false). Each subject received a series of 32 sentence/question pairs--four in each treatment combination. The noun serving as agent of each sentence was counterbalanced across subjects. Thus, each pair of subjects received the same set of stimuli but with the agent and object reversed in each sentence. Across pairs of subjects order of presentation and matching of sentences to treatment combinations was individually randomized.

Apparatus. The eye movement data were collected using a Biometrics Model SC limbus reflection eye movement monitor (Young & Sheena, 1975) interfaced to a DEC PDP-11/40 laboratory computer. Stimuli were displayed under computer control on a DEC VT-11 CRT with hardware character generator and P-4 phosphor. A more complete description of this apparatus can be found in McConkie, Zola, Wolverton, & Burns (1975).

Horizontal eye position was sampled at a rate of 1000

times/sec. Stimuli were presented one line at a time and vertical eye movement measures were not recorded. Head movement was minimized by means of a headreast and a bitebar. The latter was molded in dental wax individually for each subject. The data were stored on-line in a disk file for subsequent analysis. Also interfaced to the computer was a small subject response device consisting of an array of microswitches used by the subjects to advance the lines of text and indicate responses during the verification task.

Procedure. At the start of the experimental session each subject was asked to read a brief story while being monitored by the eye movement apparatus and then to take a brief short-answer test about what was read. This task served to refamiliarize the subjects with the apparatus and procedures of the laboratory.

Additionally, most participants read several other passages related to other research. In no case did these tasks require the subjects to do more than simply read brief passages and answer questions about them.

The sentence verification task was then described to the subjects and they were told to respond as rapidly as possible without making errors. They were then given the response unit which they held in both hands such that their thumbs rested on the "true" and "false" buttons. A third button between these two served as a "continue" button.

After a brief calibration procedure, detailed

instructions for the task were displayed on the CRT. These instructions described the sequence of events in each trial and the use of the buttons for advancing the display and responding. Each subject responded "True" with the right hand button and "False" with the left hand button.

Each trial consisted of the following sequence of events: (1) a fixation point appeared at the far right hand corner of the display area. The subject fixated this point and pressed the "continue" button. (2) The target sentence was displayed starting 10 character positions from the left hand edge of the display area. The subject read the sentence and pressed the "continue" button. (3) Another fixation point was displayed, this time at the far left hand edge of the display. The subject fixated the point and pressed "continue". (4) The probe sentence was displayed, also 10 character positions to the right of the leftmost edge of the display. The subject read the sentence and responded "True" or "False" by pressing the apporpriate button. (5) The subject pressed "continue" to begin the next trial. All 32 practice trials were performed without interruption. Subjects were then given a short break, followed by reaclibration and the 32 experimental trials. the entire procedure took approximately 20-30 minutes per subject.

In addition to the eye movement record, the computer recorded the times at which each button was pressed. This information allowed subsequent computation of response times

to a resolution of 1 msec and also allowed responses to be ralated precisely to recorded eye position.

Results

Reaction time data. The RT data were analyzed via a 2x2x2 within subjects analysis of variance. The three factors were sentence voice (active/passive), probe voice (active/passive), and correct response (true/false). Each subject contributed one observation per cell consisting of the mean RT of all correct responses in that condition. RTs were measured from the beginning of the first fixation after stimulus onset. Error rates were positively correlated with RTs, averaging 5% in the voice match conditions and 14% in the mismatch conditions.

As expected, sentence voice and probe voice interacted strongly, F (1,24) = 22.65, $p < .001$, such that RTs were faster when the voices matched. This replicates the familiar Voice Matching effect typically found in this task. The three-way interaction was also significant, F (1,24) = 4.91, $p < .04$. As seen in Figure 1, this interaction results from a generally smaller magnitude voice matching effect for False responses. RTs on False trials were also considerably more variable than on True trials. These results suggest that subjects were quite variable in the strategies and/or criteria they applied for producing False responses. For this reason, the subsequent analyses will focus mainly on True trials.

Eye movement data. The primary purpose of this analysis is to determine at what stage or stages of the verfication process the extra response time on voice mismatch trials was distributed. For this purpose, the basic dependent measure to be employed is "gaze duration" (Just and Carpenter, 1976; 1978). This measure is derived by summing the total duration of all fixations within a given text unit. The unit chosen for analysis is the sentence constituent, i.e., the subject noun phrase, the verb phrase, and the object noun phrase. Thus, each fixation was classified as falling into one of five zones of each displayed line: zone 1--before the first NP; zone 2--within the first NP; zone 3--within the VP; zone 4--within the second NP; zone 5--after the second NP. The boundary between zones was considered to be the middle of the space between the respective constituents. Since each constituent was exactly 10 characters long (except the passive VP's), zones 2 and 4 were in all cases exactly 11 character positions long, as was zone 3 for active sentences. For passive sentences, zone 3 was 18 character positions long. Comparison across treatment conditions would be facilitated if this difference in length between active and passive sentences could be adjusted for. One obvious approach is to proportionately adjust the gaze durations of the passive VP data so as to "standardize" them to a length of 11 characters. In order to obtain some estimate of the appropriateness of such an adjustment, the mean gaze

duration of zone 3 of the target (not probe) sentences was computed seperately for the active and passive sentences. Mean VP gaze duration for active sentences was 807 msec., for passives the value was 1262 msec. Thus, the ratio of active to passive times was .64. This compares well to the expected value of .61 obtained by taking the ratio of VP letter lengths (11:18). On this basis, the passive VP letter length adjustment has been adopted in most subsequent analyses.

Figure 2 shows mean total forward and regressive fixation time in each zone for voice match and voice mismatch probe sentences. An examination of this figure indicates that, overall, more time was spent fixating both the VP and the second NP than the first NP. Furthermore, these two constituents also showed the greatest increase in fixation time in the mismatch conditions, with the largest single increase occurring in forward fixation time on the second NP (averaging 219 msec). Both of these findings are compatible with the ILA hypothesis. It has previously been shown that readers tend to spend more time fixating on the main VP than on other areas of sentences (Wanet, 1971). Under the assumption of ILA, the comparitavely long times spent in the second NP would reflect post ILA decision processes. This is further supported by the finding that the bulk of the additional time resulting from voice mismatch occurs during fixations at the second NP. These findings suggest that the greater part of the voice matching

effect is attributable to processes occuring after the initial reading of the sentence.

On the other hand, there is some indication of a voice-mactching effect in all constituents, suggesting the possibility that mismatch may have had an effect on the initial sentence encoding as well as on the subsequent decision processes. The data as presented in Figure 2 are inadequate to decide this point for several reasons. First of all, no distinction has been made between the initial reading of a constituent and subsequent re-readings. Also, there is no way to distinguish between regressions within a constituent and those launched from another zone. Thus, for example, it may be that the the additional time spent in the first NP on mismatch may be due to spontaneous regressions during a post ILA symbol matching process.

To eliminate these problems, an additional analysis was performed in which mean gaze duration in each zone was classified according to the zone which immediately preceded it. That is, total mean gaze duration was tabulated separately for the time after the transition from each zone to each other zone. This procedure resulted in the 6x6 transition matrices shown in Table 1. The columns of these matrices represent the zone before each transition. Column 1 (labled "start") represents the first zone fixated after stimulus onset. The rows represent the zone after the transition and the entries show the total mean gaze duration after each transition. An exception to this is row number 1

(labeled "finish") which contains the gaze durations for the final zone before the subjects' responses. These final gaze durations have been excluded from the other entries of the table under the assumption that they contain the bulk of the final decision times. In an effort to further distinguish initial reading time from subsequent processes, the data were further divided into two groups, labeled "initial read" and "reread". The initial read group contains all the fixation time up to the first regression past the zone 4/zone 3 boundary. That is, all fixations are considered initial read until the reader first inspects the second NP (or the "after" zone) and then regresses to a previous zone. All subsequent fixations up to the response are included in the reread group. Table 1 contains four transition matrices, showing initial and reread times for the match and the mismatch conditions, respectively.

Figure 3 shows mean gaze duration on each constituent separately for initial and reread and for match and mismatch. These values are obtainable by summing the rows of the respective matrces in Table 1. Also shown are the corresponding times with final fixation time included (i.e., with the appropriate entry from row #1 of the matrix added in). The figure shows a consistently higher gaze time on initial reading for both the first NP and the VP. The near zero values for initial reading of the second NP are an artifact of our definition of "initial read" and the fact that final zone before response has been excluded. The only

data that can contribute to these values are initial regressions from zone 5 (after second NP) to the second NP. Thus, the four points consisting of match and mismatch at first NP and VP constitute the critical (but rather conservative) test of the hypothesis that mismatch affects initial encoding processes. This hypothesis was tested via a 2-way ANOVA using the appropriate entries from individual subject transition matrices as data. This analysis showed a significant effect of zone, $F(1,24) = 23.53$, $p < .001$, and of match/mismatch, $F(1,24) = 6.042$, $p < .03$. The interaction did not approach significance. Thus, even when cautiously excluding data which may be contaminated with decision time, there is clear evidence that voice match between target and probe facilitated the initial encoding of the probe sentence.

Discussion

The data presented here indicate that the majority of the added response time produced by voice mismatch in the present task is attributable to decision processes occuring after the initial reading of the sentence. The relatively long times typically spent on the second NP of the probe sentences seem to reflect post-perceptual cognitive manipulations of the semantic information from the two sentences. This result may be considered as support for a weak of the ILA hypothesis, in which the sentence is read and encoded into some unspecified internal representation prior to the comparison process. However, it is also quite

clear that this initial encoding of the sentence was not unaffected by the experimental manipulation. A relatively small but clear component of the total difference in RT occurred during the initial reading of the probe sentence.

This result argues against the strongest form of the ILA position, namely that the initial perception of the sentence occurs automatically and without regard to the semantic content of the sentence or the use to which it will subsequently be put. In the present task, a surface grammatical match between target and probe facilitated the initial reading of the probes, as well as the subsequent decision processes. This clearly suggests a more active expectency driven process than is implied by an extreme form of ILA.

GENERAL DSCUSSION

These results point toward a two-stage model of at least this particular reading task. An active primary lingusitic encoding stage transforms the visual stimulus into some initial internal representation which is stored in a buffer store. Here information accumulates until an appropriate or convienent time arises for further, more semantic processing. In the present case, such a time would be the end of the sentence. In a more normal reading situation, perhaps phrase or clause boundaries might serve as integrating points (Jarvella, 1971; Sachs, 1967).

It is important to characterize the nature of the

representation of the information that results from the initial perceptual stage. The fact that a strong voice matching effect occurs at a stage which presumably accepts this code as input argues strongly that the surface form of the sentence in preserved in the buffer. While it is possible to imagine an abstract propositional code which preserves such information, perhaps the most parsimonious approach at this point is to side with those who favor an accoustic (or articulatory) recoding view of reading. If one grants the existence of a generalized accoustic/articulatory store (c.f., Baddeley's, 1976, articulatory rehearsal loop), it is plausible that accoustic recoding during at least some reading tasks takes place not because it is a necessary stage under all circumstances, but precisely because it allows the use of such a store to smooth the comprehension process.

It should be emphasized at this point that nothing in the present data necessarily implicate such a two-stage process during "normal" reading. The case appears quite compelling for the present, rather pecular task, but the extent to which it can be extended to other task domains is an empirical question.

Aside from being rather vague, the theoretical viewpoint sketched above is also not particularly original. The primary point to be made in discussing it in the context of the present research is that the experimental techniques employed here suggest the possibility of operationalizing

such a model to a degree not before possible. The concurrent measures of processing time provided by the eye movement data have provided a quite direct method of empirically distinguishing our two hypothetically distinct stages.

There have been mixed opinions among reading researchers as to the usefulness of eye movement data in studying higher language processes. On one hand, it has proven difficult to demonstrate the effects of higher-level text characteristics on eye movement records; a fact which has lead some to conclude that the eye is only loosly yoked to cognitive activities. On the other hand, it has become clear that many kinds of low level manipulations of text can have demonstrable effects on the very next fixation (Raynor, 1975; McConkie, 19xx). Such apparently contradictory findings may be understandable within a framework such as the one suggested here. Any given experimental manipulation could have its principle effect on either the initial perceptual encoding, on subsequent semantic processes, or--as in the present case--on both. Low level manipulations such as misspellings, reduced peripheral information, etc. would be expected to affect initial encoding and should be immediately observable. More linguistic variables such as the presence of semantic anomolies or the pragmatics of functional sentence perspective (Grice, 1967; Halliday and Hasan, 197x) may well show their effects only after some critical constituent boundary has passed or some other

processing event has occured. The latter class of variables may in some cases be expected to require rather sophisticated analyses and explicit theories to make their effects observable in eye movement records. Specific task demands, such as a requirement of speeded responses and the like might also prove important in predicting the effects of various semantic manipulations. A great virtue of eye movement data is that they provides valuable constriants on theories of such phenomena. Too often, the psycholinguist is in the situation of proposing a necessarily complex model of a complex task that can only be tested against a few data points derived from mean reaction times. The kinds of techniques employed in the present research hold some promise of improving the ratio of data points to experimental parameters in this area of language theory.

References

Anderson, J. R. Verbatim and propositional representation of sentences in immediate and long-term memory. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 149-162.

Anderson, J. R. Language, memory, and thought. Hillsdale, N.J., Lawrence Erlbaum, 1976.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory--A constructive versus interpretive approach. Cognitive Psychology, 1972, 3, 193-209.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Carpenter, P. A. & Just, M. A. Reading comprehension as eyes see it. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J., Lawrence Erlbaum, 1977.

Clark, H. H. Linguistic processes in deductive reasoning. Psychological Review, 1969, 76, 387-404.

Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Clark, H. H. & Haviland, S. E. Comprehension and the given-new contract. In R. O. Freedle (Ed.), Discourse processes: Advances in research and theory (vol. 1). Norwood, N.J.: Ablex, 1977.

Clements, P. The effects of staging on recall from prose. Unpublished doctoral dissertation, Cornell University, 1976.

Fodor, J. A., Bever, T. G., & Garrett, M. E. The psychology of language. New York: McGraw Hill, 1974.

Garrod, S., & Trabasso, T. A dual memory information processing interpretation of sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 155-167.

Glucksburg, S., Trabasso, T., & Wald, J. Linguistic structures and mental operations. Cognitive Psychology, 1973, 5, 338-370.

Grice, H. P. Logic and conversation. William James Lectures, 1967, Harvard University, Cambridge, Mass.

Haber, R. N. Control of eye movements during reading. In R. A. Monty & J. W. Senders (Eds.), Eye movements and psychological processes. Hillsdale, N.J., Lawrence Erlbaum, 1976.

Just, M. A. & Carpenter, P. A. Eye fixations and cognitive processes. Cognitive psychology, 1976, 8, 441

Levelt, W. J. M. Studies in sentence perception: 1970-1976. Mimeo, Max-Planck Projektgruppe fur Psycholinguistik and Nijmegen University, 1977.

McCawley, J. D. The role of semantics in a grammar. In E. Bach & R. T. Harms (Eds.), Universals in lingusitic theory. New York: Holt, Rinehart, and Winston, 1968.

McConkie, G. W. & Rayner, K. The span of the effective stimulus during fixations in reading. Perception and psychophysics, 1975, 17, 578-586.

McConkie, G. W. & Rayner, K. Asymmetry of the perceptual span in reading. Bulletin of the Psychonomic Society, 1976, 8, 365-368.

Mehler, J., Bever, T. G., & Carey, P. What we look at when we read. Perception and Psychophysics, 1967, 2, 213-218.

Olson, D. R. Language and thought: Aspects of a cognitive theory of semantics. Psychological Review, 1970, 77, 257-265.

Olson, D. R. Language use for communicating, instructing, and thinking. In R. Freedle & J. Carroll (Eds.), Language comprehension and the acquisition of knowledge. Washington, D.C.: V H Winston, 1972.

Olson, D. R., & Filby, N. On the comprehension of active and passive sentences. Cognitive Psychology, 1972, 3, 361-381.

O'Regan, J. K. Structural and contextual constraints on eye movements in reading. Unpublished doctoral dissertation, University of Cambridge, 1975.

Perfetti, C. A & Goldman, S. R. Discourse functions of thematization and topicalization. Journal of psycholinguistic research, 1975, 4, 257-271.

Rayner, K. Visual attention in reading: Eye movements reflect cognitive processes. Memory and Cognition, 1977, 5, 443-448.

Rayner, K. & McConkie, G. W. What guides a reader's eye movements? Vision research, 1976, 16, 829-837.

Schank, R. C. Conceptual information processing. Amsterdam: North-Holland, 1975.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Tanenhaus, M. K., Carroll, J. M., & Bever, T. G. Sentence-picture verification models as theories of sentence comprehension: A critique of Carpenter and Just. Psychological review, 1976, 83, 310-317.

Tinker, M. A. Recent studies of eye movements in reading. Psychologiccal bulletin, 1958, 55, 215-231.

Trabasso, T. Mental operations in linguistic comprehension. In R. Freedle and J. Carroll (Eds.), Language comprehension and the acquisition of knowledge. Washington, D.C.: V H Winston, 1972.

Wanat, S. F. Lingusitic structure and visual attention in reading. Newark, Deleware: International Reading Association, 1971.

Winograd, T. Understanding Natural Language. New York: Academic Press, 1972.

Table 1 -- Mean initial and re-read transition matrices for match and mismatch conditions. Columns and rows represent zones before and after transition, respectively. Entries are mean fixation times (in msec.) after transition (except "finish" rows, which show times in last zone before response).

MATCH

Initial

	Start	Before	NP 1	VP	NP 2	After	Total
Finish	0.	0.	8.	66.	264.	11.	349.
Before	29.	0.	1.	0.	0.	0.	30.
NP 1	158.	28.	0.	15.	0.	0.	201.
VP	89.	5.	205.	0.	0.	0.	299.
NP 2	0.	0.	0.	10.	0.	4.	15.
After	0.	0.	0.	1.	2.	0.	3.

Reread

Finish	0.	0.	9.	29.	23.	0.	61.
Before	0.	0.	0.	0.	0.	0.	0.
NP 1	0.	0.	0.	3.	8.	0.	11.
VP	0.	0.	1.	0.	18.	0.	19.
NP 2	0.	0.	0.	91.	0.	0.	91.
After	0.	0.	0.	0.	1.	0.	1.

MISMATCH

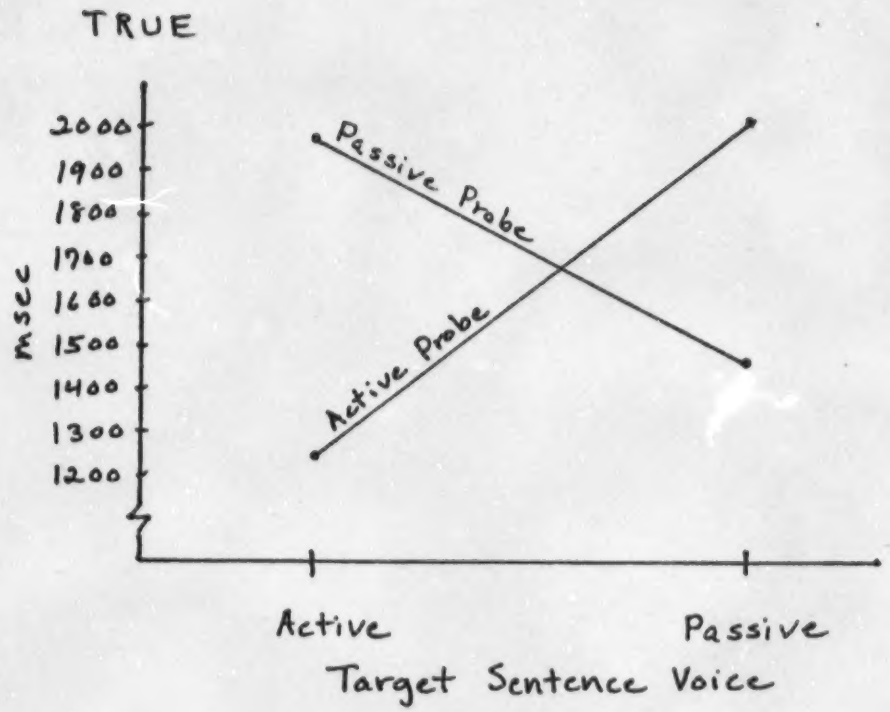
Initial

Finish	0.	0.	19.	76.	352.	18.	465.
Before	13.	0.	6.	3.	0.	0.	22.
NP 1	192.	19.	0.	24.	0.	0.	235.
VP	93.	6.	254.	0.	0.	0.	353.
NP 2	0.	0.	7.	11.	0.	7.	25.
After	0.	0.	0.	2.	7.	0.	10.

Reread

Finish	0.	0.	34.	66.	55.	0.	155.
Before	0.	0.	0.	0.	0.	0.	0.
NP 1	0.	0.	0.	10.	18.	0.	28.
VP	0.	0.	21.	0.	71.	0.	92.
NP 2	0.	0.	4.	203.	0.	0.	207.
After	0.	0.	0.	0.	0.	0.	0.

Figure 1 - Mean Reaction Time For Correct Responses In Each Condition



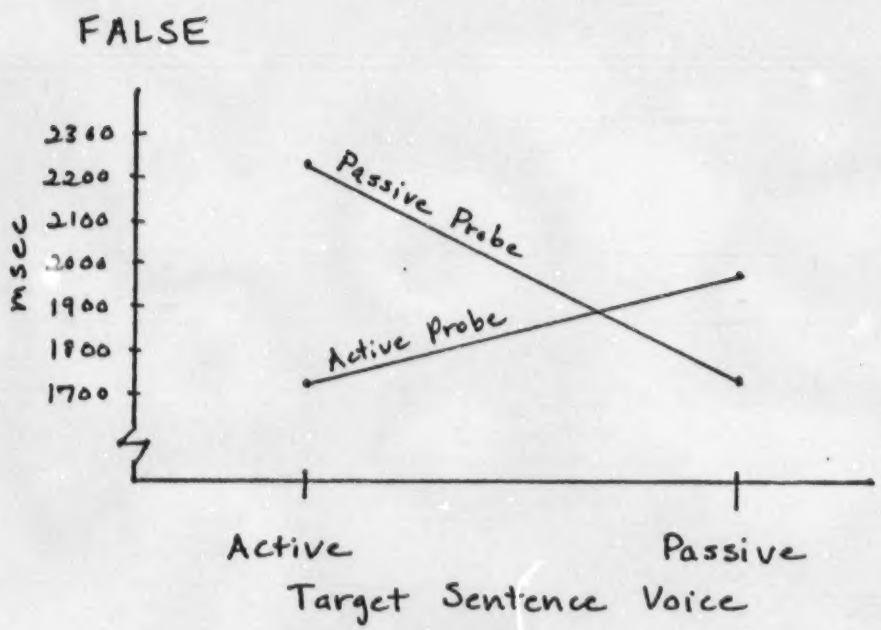


Figure 2 - Mean Total Fixation Time on True Trials in each Sentence Zone

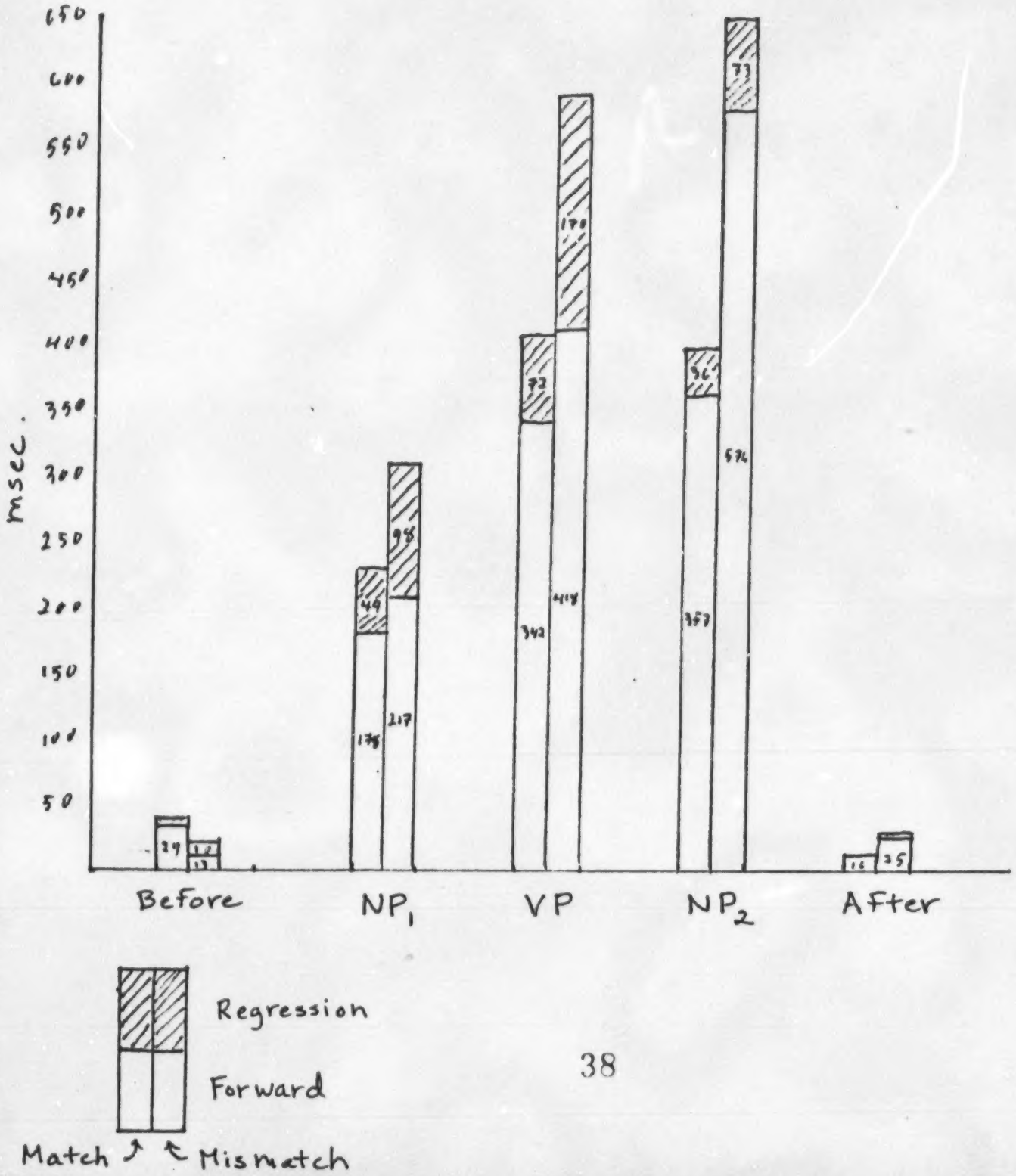


Figure 3 - Mean Gaze Duration in each Constituent For Initial Reading and Re-reading (True Trials)